Answer to Anonymous Referee #1 comments

We thank the Referee #1 for the valuable comments which we used to revise our manuscript. In the following we state each referee comments in italics and answer directly below to each of them also highlighting how and where we changed the manuscript accordingly. All changes in the manuscript are visible in red, except for Table 2 because here red signifies already "medium quality".

Note:

1) During revision we noticed that 5 of our samples in the satellite data set have not been used by mistake in the model due to some annotation error in the input data. We now revised the input data description and by that increased the number of sample for the 1x1 (now 139 instead of 135), 3x3 (now 155 instead of 150) and 5x5 (now 160 instead of 155) collocations. Results are still very similar, but because of that all related Figures (Fig. 6, 7d-f and 8-according to new figure names) and Tables (Table 1, Table 2b, Table 3c, Table 4) are slightly changed.
2) As pointed out by Referee #2 we have used by mistake the wrong PPC (=Allo+Diadino+Diato+Viola+Lut+Zea+Caro) definition for Hooker et al. (2005) with and now revised it to the correct one (=Allo+Diadino+Diato+Zea +Caro), accordingly for all models the PPC results were recalculated and all related Tables (in addition to the above mentioned (Table 2a, 3a, 3b and Fig. 5c, 6a-c and 8f-according to new figure names) were revised.

General comment This study reports the development of an algorithm based on empirical orthogonal functions (EOFs) for detecting phytoplankton pigments using field- and satellite-measured spectra of remote sensing reflectance. This is a methodological paper and it is appropriate for Ocean Science. Developing new methods for deriving phytoplankton pigments from space is very interesting for remote sensing community.
Detection of individual pigments or pigments groups is not useful only for the retrieval of phytoplankton functional types, which is still a debate, but also for examining the physiology of phytoplankton over a wide scale. I appreciated the use of EOF analysis which is a robust method of investigation, as well its use in oceanic waters, for which this kind of applications are always problematic. The models are well developed and internally validated. However, the model validation made by using a totally independent data set of pigment data is missing. Validation is an important step to understand if the method can really be applied. However, a comparison with other approaches and an application over a wide area using satellite data are present and they are useful to reinforce somehow model results. I think the paper is worth to be published but there are several aspects that need further clarification before publication (see list below).
In addition, I found the manuscript difficult to follow and sometimes confusing. I suggest revision of the english.

We thank the reviewer #1 for his/her valid suggestions. In order to improve the structure of our manuscript we have added as suggested a flowchart (new Fig. 2) giving an overview on our new method. We reply in the following to all the points asked for clarification which were also incorporated in the revised manuscript. Our revised manuscript had a final revision of the English by our U. S. American coauthor.

The models are well developed and internally validated. However, the model validation made by using a totally independent data set of pigment data is missing. Validation is an important step to understand if the method can really be applied. However, a comparison with other approaches and an application over a wide area using satellite data are present and they are useful to reinforce somehow model results.

In order to test our EOF methods for independent data sets using the method established by a certain testing data set, we have used the cross validation technique. The technique allows resampling the whole data for 500 different subsets (i.e. run by 500 permutations) into testing and validation data sets which are independent from each other. To discuss limits and application of our method further we have added this explanation to the discussion (end of section 3.5).

Specific comments:
Title: I suggest ". . ..for the prediction of phytoplankton pigment concentrations".
Was changed accordingly.

Page 2076, lines 15-20: PSC are also involved in light harvesting like accessory chlorophylls.
Actually this is exactly what we have written here: "Besides Chl *a* there are many other pigments in phytoplankton that are either involved in light harvesting (such as chlorophyll *b* (Chl *b*), chlorophyll *c* (Chl *c*) and several carotenoids, called photosynthetic carotenoids (PSC)), or protecting Chl *a* and other sensitive pigments from photodamage (photoprotective carotenoids, PPC)."

Page 2076, line 20: a reference which summarizes the distribution of taxonomic pigment among the algal classes could be useful. I don't think Mackey et al. (1996) is appropriate here.
We changed accordingly the reference to Letelier et al. (1993) and Vidussi et al. (2001) and removed the reference to Mackey et al. (1996).
Letelier, R. M., Bidigare, R. R., Hebel, D. V., Ondrusek, M., Winn, C. D., Karl, D. M.: Temporal variability of phytoplankton community structure based on pigment analysis. Limnol. Oceanogr., 38(7), 1420-1437, 1993.

Vidussi, F., Claustre, H., Manca, B.B., Lucheta A. and Marty J.-C.: Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter., Journal of Geophysical Research, 106, 19939–19956, 2001.

Page 2076, line 26: what do you mean for "overall biomass"? Generally Tchl a is considered a proxy of phytoplankton biomass. Please, reformulate the sentence.
We reformulated the sentence to: "TChl *a* is far from being a sole function of phytoplankton biomass and varies as other phytoplankton pigments well with taxonomic composition and mean physiological state of the algal assemblage in response to several factors, e.g., light, temperature, and nutrients (Behrenfeld and Boss 2006)."
Behrenfeld, M. I., and Boss, E. B.: Beam attenuation and chlorophyll concentration as alternative optical indices of phytoplankton biomass. Journal of Marine Research, 64, 431-451, 26.

Page 2077, lines 4-7: it is not clear if these studies are referring just to the surface layer or to the water column.
To clarify we changed the sentence to "Several r ecent studies have investigated the potential of using continuous optical data to derive surface concentrations of pigments other than TChl *a*,

with the advantage of being able to supply estimates of larger spatial and temporal scale than obtained with *in situ* water sampling."

Page 2077, lines 15-18: I think it is not only a problem of atmospheric correction but also of appropriate inversion models. If AOPs-inversion models are accurately validated, also inherent optical properties could be used to derive pigment composition.
We agree and therefore we changed the sentence to "However, a_p and a_{ph} are inherent optical properties (IOP) which cannot be directly determined from satellite ocean colour measurements after successful atmospheric correction, such as the apparent optical properties (AOP). The estimation of IOP from AOP is based on a certain inversion model (e. g., the Quasi-Analytical-Algorithm by Lee et al. (2002), which introduces additional uncertainty."
Lee, Z. P., Carder, K. L, and Arnone, R. A.: Deriving inherent optical properties from water color: A multi-band quasi-analytical algorithm for optically deep waters, Applied Optics, 41, 5755-5772, 2002.

Page 2078, lines 12-20: I suggest to make the aim of the study clearer.
We have re-formulated the aim to "The EOF analysis is applied to R_{RS} and to $\rho_{\_wN}$ (i.e. normalized ρ_w just above surface) data measured in the field and by satellite sensors, respectively, in the Atlantic Ocean. The dominant EOF loadings were subsequently assessed as predictors in a multiple linear regression for the concentration of phytoplankton pigments and pigment groups as response variables. The prediction error of each model is evaluated by a permuated cross-validation routine, which is used to estimate the critical sample sizes necessary for reliable prediction. In addition, we demonstrate the approach's utility in estimating the large scale distribution and photo-physiology of the phytoplankton assemblage."

Section 2.2: I suggest to insert any source of pigment data you used or to add this information in Table S1.
We do not understand this comment, all sources of pigment data we used here are listed in Table S1.

Section 2.2: In addition, why did you use also 3X3 and 5X5 matchups? Along the text, you sparingly speak about them. Some results are just shown in table 4.
We tested if it is critical to use the closest collocations or if it is better to have a representative mean value. Results showed that for the pigment predictions and the validation of TChl *a* Polymer product this was not critical. We wrote briefly about the results of all three type predictions in the first paragraph of 3.3.2. Since our motivation was not clear we added here "Results showed that even models based on 5x5 pixel collocations can produce robust results."

Page 2082, line 9: "different combinations. . ..", please be more clear.
To clarify we change the whole first paragraph of 2.3 to "Fig.1 presents the distribution of collocated pigment and reflectance measurements for both, field and satellite-based, data sets which were used separately as input for the EOF prediction analysis. The field data set covered 53 collocated reflectance and pigment data points (Fig. 1, red points). We used three set-ups of the field $R_{RS}(\lambda)$ spectra for the development of pigment specific models:

a. $R_{RS}(\lambda)$ data in hyperspectral (1 nm resolved, "hyper_R_{RS}") from 350 to 700 nm,

b. "hyper_R_{RS}" from 380 to 700 nm, and
c. $R_{RS}(\lambda)$ data in MERIS band resolution ("band_R_{RS}").

The three satellite-based data sets consisted of 139, 155 and 160 collocated reflectance and pigment data points from 2002 to 2012 for the 1x1 (Fig. 1 stars), 3x3 (Fig. 1 diamonds) and 5x5 (Fig. 1 squares) pixel collocation criterions, respectively, covering all months except January, March and December (details on the spatial and temporal distribution of collocations are given in the Supplement Table 1). 18 collocations of the field data matched the 1x1 pixel satellite-based data set (Fig.1, red stars), but none matched the other two satellite-based data sets."

Various pages in the methods section: A flowchart describing the various steps of your algorithm would be very useful to the reader.
We have now introduced as a new figure (Fig. 2) the requested flowchart and added in section 2.3.1 the following text "Fig. 2 gives an overview describing the various steps of the development and validation of our EOF method to predict various pigments and pigment groups' concentrations which are described in details in the following subsections."

Section 2.3.1: How many models did you develop? Because of the two spectral ranges for hyperspectral Rrs (350-700 nm and 380-700 nm), are two different models produced? It is not clear.
In fact, we have tried a lot of different setups for the hyperspectral data set (350 to 700nm, 380 to 700 nm, 400 to 700 nm, 400 to 550 nm, 300 to 800nm). As now explained in the flow chart (new Fig. 2) we focus for the hyperspectral models on the 350 to 700 nm and 380 to 700 nm only. We have also clarified that in the text in section 2.3.1 1st paragraph (methods), 3.3.1 1st paragraph (results) and show in detail both models' results now in the Supplement Table 3.

Please, explain also why you normalized the spectra. I expect that pigment concentrations are more related to the magnitude.
As it has been shown in Craig et al. (2012) if the spectra are not normalized than the first EOF mode just reflects the coarse variation in the water's colour, and is likely the signature of bulk oscillations in biomass concentration but does not reveal spectral changes caused by different specific pigment signatures. EOF-2, EOF-3 and EOF-4 modes in Craig et al. (2012) reflected than the same spectral shape as the EOF-1, EOF-2 and EOf-3, respectively, of the standardized spectra. With the same success, the EOF analysis has been used in Taylor et al. (2013) to retrieve PE conc. which we explained in the method section 2.3.1 1st paragraph ("As in Taylor et al. (2013), spectral datasets **X** were standardized for each sample row by first subtracting the mean spectral value (centering) followed by division by the spectral standard deviation (scaling), which focused the analysis on the spectral shape rather than the magnitude.") .

Page 2085, eq 4: It is not necessary r-squared formula. Page 2085, eq 5-8: As you calculated other statistics using log-transformed quantities, why are RMSE, MPD, PB and MDPD calculated for no-log quantities? I suggest to use same quantities.

As suggested we have deleted the r-square formula. We have calculated the statistics partly on log-tranformed (R^2, R^2cv) and non log-transformed data (all other parameters: RMSE, MPD, PB,

MDPD, RMSEcv, MPDcv, MDPDcv) because we wanted to have statistics developed in the same matter as the studies we were comparing our results to:

a) Craig et al. shows results for full-fit and cv (for TChl *a* only) with RMSE and R^2 both on log scale -no statistics on MPD, MDPD was shown- only absolute bias which we think when looking at several pigments is not conclusive enough.
b) Chase et al. shows results using a different method (Gaussian) on IOP data with MDPD (median percent difference) only on normal (i. e. non-log) scale.
c) Pan et al. shows statistics for fitting and validation with RMSE and R^2 on log-scale while MPD and Bias are shown on normal scale.
d) Brewin et al. 2014 showed the validation of the OCV4 Tchla algorithm based on non-log satistics.

To our understanding the median percent error really only makes sense for the non-log transformed data because the error will increase proportionally with the values (i.e. as a percent). This is actually the underlying assumption when the model is fit to log-transformed data. However, the minimization of error in the fitting of the model is indeed based on the RMSE of the log-transformed data. Therefore, thanks to your and Referee #2 comments we have noticed that we should provide the statistics on RMSE on log-scale. Now we changed our comparison of predicted versus observed pigment concentrations to RMSE and RMSEcv calculated on log-tranformed data. Accordingly we revised equations #4 and #13, and recalculated RMSE and RMSEcv values for all pigment and pigment group concentrations from all models. Accordingly Table 2, Table 4, Fig. 5, Fig. 6, Fig. 7c and Fig. 7f (according to new figure names) were modified. For Fig. 7c and 7f we present now the jack-knifing results for RMSEcv as ratio to the RMSE (as it was done for R^2 and MPD). In summary, the recalculation of RMSE and RMSEcv on log-trabsformed comparisons still proofs the validity of our method. In order to still compare our EOF method to retrieve TChl *a* from MERIS reflectance data to the global OC4V4 validation results by Brewin et al. (2014) we now changed the text in the 3rd last paragraph of section 3.5 by referring to our results in the BGD paper ("Global validation by Brewin et al. (2014), with 1039 collocations and retrievals of TChl *a* directly from *in-situ* $\rho_{wN}(\lambda)$ data, showed for OC4V6 an R^2 of 0.87 and a *RMSE* of 0.29 based on non-log statistics (which compares to our RMSE values on log-scale shown in Table 4 of Bracher et al. 2014).")

Section 2.3.4: it is not necessary or I suggest to move it in another section.
This subsection is actually necessary to explain how we obtained the pigment predictions from one month of MERIS reflectance data. To clarify we changed the subsection to "In order to predict pigment concentration from MERIS $\rho_{wN}(\lambda)$, for a whole month of data in November 2008, where we did not have corresponding pigment measurements, the following method was applied: we projected standardized MERIS $\rho_{wN}(\lambda)$ data onto the EOF loading (**V**) to derive their principal components (**U**), which were subsequently used for the prediction with the fitted linear model (as in Sect. 2.3.3, step7, eq. 11, Fig. 2, right panel), where $b_{1,2,...n}$ are taken from the EOF model developed with the 1x1 MERIS Polymer $\rho_{wN}(\lambda)$ matchups (following Fig. 2, left panel)."

Section 2.4: could be included in section 2.2.
Was changed, as suggested.

Page 2088, lines 15-20: They are not results and could be included into the methods.
As suggested, was moved to the beginning of section 2.3.

Page 2089, lines 3-16: This paragraph is confused and contradictory. You say that range of pigments is similar between the two datasets and then that the maxima and minima of one dataset are higher than those of the other. The paragraph has to be reformulated.
As suggested, we reformulated the paragraph to: "The composition and range of pigments (as detailed with maximum, minimum, mean and standard deviation in Supplement -Table 2) shows for all pigments (except for Fuco for which it is equal, and for Zea for which it is inversed) that the collocations to the field data set contain higher maxima and minima than the collocations to the satellite-based data set. Mean values are for most pigments very similar among both data sets, but standard deviation is rather 2-3 times higher than the mean value for all pigments (except for PPC and Zea) in the field data set as opposed to the standard deviation being of a magnitude similar to the mean value in pigment data collocated to the satellite data set. However, DVChl *b*, MVChl *b*, TChl *b*, Allo, Diato, Lut, Neo, Peri, Viola and TPheo had values of 0 mg m^{-3} in more than 20% of all stations in both data sets. Also Chl-c_3 had a concentration of 0 mg m^{-3} in one sample collocated to the field and in over 30% of samples collocated to the satellite-based data set. Several pigments had concentrations of 0 mg m^{-3} only occasionally (<10%) in samples collocated to the satellite-based data set (Caro, Chl $c_{1/2}$, But, Hex, Zea, DVChl *a*, Diadino and Fuco), and in the field data sets (DVChl *a*, Diadino and Fuco). All other pigments not listed here had detectable concentrations in all samples."

Page 2089, line 27 to the end of the sentence: "Still one has to bear in mind. . .." You did not explain it in the method section.
We deleted this sentence now because it is not really necessary to be mentioned here. But we have now, in response to Referees #2 comment, briefly described the measurement in chapter 2.1.2:
"For all three cruises as AOP input data, we used $R_{RS}(\lambda)$ data obtained from profiles of radiance and irradiance from 320 nm to 950 nm with an optical resolution of 3.3 nm and a spectral accuracy of 0.3 nm measured with hyperspectral radiometers (RAMSES, TriOS GmbH, Germany) measuring at the same time and place when pigment data of section 2.1.1 were sampled."

Page 2090, line 25: "more or less" is not appropriate.
We agree and it was deleted!

Page 2091, lines 1-4: "This indicates that. . ...", I think this is a conclusion that you can draw after description of also EOF-4 which is related to the different pigment composition.
Was moved below as suggested.

Page 2091, lines 6-8: you say that Rrs amplitude is affected only by pigment absorption. What about backscattering influence?
We suspect rather EOF-1 to be influenced by backscattering because of the less steep increase with wavelengths of EOF-1 as opposed to EOF-2. We have changed the sentence referring to EOF-1 to "EOF-1 is likely the signature of bulk oscillations in phytoplankton biomass concentration (including its effect on backscattering)."

Page 2091, lines 11-13: I don't agree with this statement. In EOF-2 the peak at 683 nm is negative, as in Craig et al. 2012, so as they suggested it is not related to chlorophyll. Only in EOF-1 you don't see the peak, because of the low chlorophyll concentrations as you said.
Was changed, as suggested.

Page 2091, lines 14-20: What about CDOM influence?
We have added CDOM here which was clearly missing!

Page 2092, line 3: "apparent" is not appropriate. You can replace with "detected", for example.
Was changed accordingly!

Section 3.3.1: You could discuss more the regressions shown in Figure 4 (now named Fig. 5).
As suggested we added a discussion of the regression in this section "For pigment groups and pigments with a high range of data (TChla, PSC and Hex), covering about three orders of magnitude, the intercept is much lower and the regression closely aligns with the 1:1 line. The predicted versus observed regression for Zea was of lower quality (R^2 <0.6) likely due to a much lower range of observed concentrations."

Section 3.3.1 The replacement of concentration of 0 mg m-3 with a small value generates only confusion.
We agree and removed that statement!

Page 2092, lines 15-17. I am not sure that bad predictions are related to the occurrence of some samples with concentrations of 0. I think that when a pigment concentration is generally low is difficult to outline its spectral shape and therefore to predict its concentration.
Well, our results for the pigments, which were only in a few samples missing, indicate clearly that by excluding these samples robust predictions were possible. However, we agree that also when the conc. of a certain pigment is low it is difficult to outline its spectral shape and therefore to predict its concentration. We added this now to the text in chapter 3.5 5th paragraph where we discuss our method's potential.

Section 3.3.2: You could discuss more the regressions shown in Figure 5 (now named Fig. 6).
As suggested we added a discussion of the regression in this section "The full-fit results shown in Fig. 6 show that the models based on the satellite data show much poorer predictions (e.g., *a*, R^2 and *RMSE*) than the field data models for all pigment or pigment groups (except Zea) even though the satellite data models are based on more samples. This may be caused by the lower quality of water leaving reflectance data obtained from satellite opposed to direct radiometric measurements in the water column. Another explanation may be that the lower standard deviation of the pigments in the satellite based data set leads to less precision of the EOF based models. The later may explain why the full-fit results for predicting Zea concentrations are very similar for the two model types."

Page 2095, line 26: "Exemplarily"?
Changed to "as examples"

Section 3.6: Prediction of pigment concentration over a wide area is an interesting application. However, as you do not have in situ pigment concentrations for validation, I suggest you to compare your pigment distribution with results from literature even if sampled in other periods.
We only found very limited information on pigment distributions of TChla, MVChla, DVChla, PPC and PSC independent of the input data (Table 1 in the Supplement) we used to construct our models for these pigments / pigment groups' predictions. We therefore only referred so far to the Longhurst Provinces and to the results of the detailed study by Taylor et al. (2011) where the same pigment measurements incorporated in our study were related to various environmental factors (light, temperature, mixed layer depth, nutrients, euphotic depth) at those specific stations. Now, we have investigated this further and only found two studies where for specific stations within the study area at different times of the year concentrations of those pigments have been measured. We added the reference to their findings at the end of section 3.6:
"Comparisons of our predictions to pigment data not used for the development and validation of our EOF model show consistent results: Partensky et al. (1996) measured in December 1992 (EUMELI 5 cruise) at a station within a phytoplankton bloom at 18°29'N and 21°05'W TChla concentrations of about 1.2 mg m^{-3}, similar to the range of our predicted values at the southern edge of the Northern bloom. Barlow et al. (2002) measured within the area of our predictions conc. of TChla, DVChla, PSC and PPC during the AMT-3 cruise in Oct 1996 at 20°N and 20°W (0.4, 0.05, 0.175 and above 0.09 mg m^{-3}, respectively) and 30°N and 22°W (0.05, 0.01, 0.022 and above 0.04 mg m^{-3}, respectively) similar to our predicted concentrations for the same pigments just east to the Northern bloom of the Mauretanian Upwelling and at the North Atlantic Tropical Gyre Province East, respectively."
New references:
Barlow, R. G., Aiken, J., Hooligan, P. M., Cummings, D. G., Maritorena, S., and Hooker, S.: Phytoplankton pigment and absorption characteristics along meridional transects in the Atlantic. Deep-Sea Res. Pt. I , 47, 637–660, 2002.
Partensky, F., Blanchot, J., Lantoine, F., Neveux, J., and Marie, D.: Vertical structure of picophytoplankton at different trophic sites of the tropical northeastern Atlantic Ocean. Deep-Sea Res. Pt. I , 43, 1191-1213, 1996.

Table 1: Add a column showing the cumulative proportion of explained variance.
The previous Table 1 does not allow a 2nd column for the cumulative proportion, therefore we added another panel showing the cumulative proportion (we do not think this is necessary as readers can easily calculate that in their head but we added it as the reviewer requested it).

Table 2: I will find useful, also for future studies and applications, if you show statistics also for bad predicted pigments.
We thought it would be too much to show all statistics, but as suggested we provide this information now in the new Table 3 of the Supplements and added appropriately the reference in sections 3.3.1 and 3.3.2.

Figure 1: A legend on the panel will be also useful.
As suggested we added a legend to the figure.

Figure 3: I suggest you to make a panel for each EOF. Please, add y-axes name.
As suggested we changed the figure (now named Fig. 4) and added the name of y-axes.

Figures 4 and 5: units for pigment concentrations are missing in the plot. Could you add also MPD, PB and MDPD and discuss them in the text?
As suggested we added the units to the plots. We think the two figures would become overloaded if also MPD, PB and MDPD were added. Therefore we now provide all the statistical parameters of the full fit (R^2, RMSE, intercept, MPD, PB and MDPD) and the mean values of validation over all permutations (R^2cv, RMSEcv, MPDcv and MDPDcv) in the new Supplement-Table 3 for all models established for all pigment and pigment groups and based on the six reflectance data sets: field data set with hyperspectral resolution (hyper_R_{RS}) a) from 350 to 700nm and b) from 380 to 700nm, c) field data set with multispectral (band_R_{RS}) resolution and satellite data set with multispectral resolution (MERIS ρ_{wN} data) using d) 1x1, e) 3x3, and f) 5x5 matchups. The two figures are now named Fig.5 and Fig. 6.

Table S2: It would be useful also information on the mean +- standard deviation for each pigment.
Added accordingly and mention and discussed these in the 2nd paragraph of Section 3.1.

Typesetting errors:
Along the text: I think it is better if you use Rrs instead of RRS.
Changed (This was indeed a typesetting error and correct in the word file submitted by us. Thank you for notifying!)- we also hanged to "Rrs"
Page 2084, line 19: please verify if the occurrence of the term "e" is appropriate.
Changed (This was indeed a typesetting error and correct in the word file submitted by us. Thank you for notifying!)

Page 2084, line 28: the intercept sometimes is called "I" and sometimes "a".
Changed accordingly

Page 2105, line 24: It is "Antoine" instead of "Antoinem".
Changed (This was indeed a typesetting error and correct in the word file submitted by us. Thank you for notifying!)